Appendix: Regression Analysis and Population Projections

Typically, life-cycle perspectives are applied in studies explaining patterns of individual-level migration behavior (see, for example, Bailey, 1993; Bailey, 1994; McHugh et al., 1995). Here, regression analysis is used to study the aggregate effects of migration decisions on different types of rural destinations. This analysis builds on previous studies that address this broader, age-cohort context in which individual life-cycle paths evolve (Plane, 1992; Plane, 1993; Pandit, 1997b; Pandit, 1997a; Nelson et al., 2004).

Multivariate regression measures the effects of county-level, socioeconomic attributes on net migration in the 1990s. Separate models by 5-year age cohorts capture the life-cycle variation of these effects. Using predicted cohort migration rates derived from the models' parameter estimates, along with data on expected death rates and immigration, this analysis projects the size and distribution of boomer migration to different types of counties through 2020. Low-, middle-, and high-range projections provide upper and lower bounds for the likely impact on nonmetro counties.

Model Specification and Variables

In these models, county-level, cohort-specific, net migration rates are set as the dependent variables. The county migration rates (CMR) were calculated for 5-year age cohorts and represent the population growth from net migration (NM) for age group a in county j during the 1990s by ages as of 2000. These rates are expressed as a percentage of the population (P) in age group a-10 in 1990:

$$CMR_{j,a,1990-2000} = \frac{NM_{j,a\ 1990-2000}}{P_{j,a-10,1990}}$$

Results are reported for 11 age groups, covering 5-year age cohorts from 25 to 79 years old (app. table 1). The models use the same set of county-level

Appendix table 1
Cohort migration rates, 1990-2000

Age range (years)	Mean	Minimum	Maximum	Standard deviation
25 to 29	-0.078	-0.909	4.773	0.320
30 to 34	0.187	-0.876	7.000	0.387
35 to 39	0.154	-0.557	2.981	0.236
40 to 44	0.112	-0.636	1.586	0.174
45 to 49	0.089	-0.755	1.263	0.145
50 to 54	0.086	-0.645	4.667	0.172
55 to 59	0.103	-0.399	1.586	0.183
60 to 64	0.107	-0.500	2.187	0.199
65 to 69	0.084	-0.545	2.130	0.175
70 to 74	0.032	-0.414	1.145	0.107
75 to 79	0.011	-0.548	0.644	0.079

Source: USDA, Economic Research Service, using data from the U.S. Census Bureau and the National Center for Health Statistics.

independent variables, measuring socioeconomic conditions as of 1990, which is the beginning of the migration period being analyzed (or 1990-93 in the case of employment change). Independent variables measure employment and housing market conditions, recreation and natural amenities, urban influence, demographic characteristics, and regional location variables representing selected census divisions (app. table 2).The endogeneity of migration and employment change is difficult to disentangle and biases modeling results if not adequately addressed. Migrants are simultaneously attracted to areas with high employment growth, but their migration stimulates further employment growth. Here, the analysis follows established methodology by using a lagged variable that measures employment change for the early 1990s to "explain" net migration measured for the entire decade (Partridge et al., 2007). Diagnostics show that this reduces endogeneity sufficiently but still captures a strong positive relationship between employment and migration among younger age groups.

Independent variables were converted to z-scores so that the measures could be expressed in relative rather than absolute terms. This is necessary because migration is a closed system—higher inmigration in one area assumes higher outmigration somewhere else. Higher unemployment in a county should not affect migration if that county's position relative to others does not change

Appendix table 2

Independent variables

	Mean	Minimum	Maximum	Standard deviation
Employment and housing market factors				
Percent unemployed, 1990	6.7	0	30.5	3.1
Employment change, 1990-93[1]	1.5	-12.8	180.3	3.9
Median home value, 1990	53,670	0	500,000	33,356
Natural amenities and recreation				
ERS Natural Amenity Index	0.1	-6.6	11.2	2.3
Percent seasonal housing units, 1990	5.7	0	75.4	9.4
Urban influence				
Urban metro, 1990[2]	0.19	0	1	0.35
Rural metro, 1990	0.08	0	1	0.28
Nonmetro adjacent, 1990	0.32	0	1	0.47
Nonmetro nonadjacent, 1990	0.41	0	1	0.49
Percent urban, 1990	36.5	0	100	29.8
Population density, 1990	226	0.1	68,157	1,705
Demographic characteristics				
Percent of married couples with no children, 1990	32.7	11.6	51.2	4.2
Percent foreign born, 1990	2.2	0	45.1	3.6
Regional location				
South Atlantic	0.19	0	1	0.38
East South Central	0.17	0	1	0.32
West South Central	0.15	0	1	0.35

[1]Average annual change in total employment, 1990-93, as a percent of 1990 total employment.
[2]In the regression analysis, this is the omitted category of the 4-tier urban-rural classification.
Note: For data descriptions, see boxes, "Data Sources" and "County Classifications Used in This Report."
Source: USDA, Economic Research Service, using data from the U.S. Census Bureau and the Bureau of Economic Analysis.

(i.e., unemployment increases in all counties). But net migration for a county would decrease if unemployment grew at a faster rate than in other counties.

Further diagnostics revealed no significant violation of assumptions of linearity, independence of the error terms, and error distribution. Variance inflation factors showed no signs of significant multicollinearity among the predictors. Scatter plots of residuals versus fitted values showed little evidence of heteroscedasticity.

Parameter estimates from variables that have been converted to z-scores allow comparison of effects across and within age-specific models (app. table 3). There is much variation by age in the overall explanatory power of the chosen set of independent variables, as reflected in the (adjusted) r-square values. The model explains around 30 percent of the variation in cohort migration rates for 25-29 and 30-34 year olds but improves to over 50 percent for 55-59 and 60-64 year olds. Migration models in general are more powerful in explaining the migration flows for those life-cycle stages through which boomers are currently passing. The high level of explanatory power also may reflect the impact of cohort size on migration “effectiveness,” the degree to which population growth from inmigration is not offset by an

Appendix table 3

Regression coefficients for cohort migration rates, 1990-2000

	Age range (years)										
Independent variables	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79
Employment and housing market factors											
Percent unemployed	**-0.145**	**-0.141**	**-0.047**	-0.024	0.011	**0.054**	**0.093**	**0.128**	**0.124**	**0.069**	**-0.048**
Employment change	**0.188**	**0.281**	**0.253**	**0.218**	**0.209**	**0.148**	**0.108**	**0.070**	**0.068**	**0.073**	**0.061**
Median home value	**0.125**	**0.088**	**0.148**	0.109	**0.052**	-0.017	**-0.041**	**-0.051**	**-0.057**	**-0.058**	**0.055**
Natural amenities and recreation											
ERS Natural Amenities Index	**0.140**	**0.174**	**0.189**	**0.229**	**0.255**	**0.282**	**0.271**	**0.261**	**0.268**	**0.216**	**0.036**
Percent seasonal housing units	0.119	**0.072**	**0.096**	**0.157**	**0.215**	**0.351**	**0.415**	**0.408**	**0.344**	**0.218**	0.027
Urban influence											
Rural metro	-0.016	0.021	**0.099**	0.094	**0.062**	**0.038**	0.019	0.015	0.032	**0.061**	**0.081**
Nonmetro adjacent	**-0.087**	**-0.083**	-0.021	-0.007	-0.004	0.005	0.032	**0.052**	**0.061**	0.040	-0.005
Nonmetro nonadjacent	**-0.173**	**-0.146**	**-0.143**	**-0.130**	**-0.133**	**-0.099**	-0.029	0.002	**0.006**	-0.029	**-0.080**
Population density	0.029	0.024	**-0.033**	**-0.033**	-0.007	0.027	**0.033**	0.034	0.013	**-0.037**	**-0.107**
Percent urban	0.027	**-0.243**	**-0.187**	**-0.129**	**-0.092**	**-0.059**	-0.012	**0.039**	**0.097**	**0.190**	**0.251**
Demographic factors											
Percent of married couples with no children	**-0.131**	**0.170**	**0.253**	**0.283**	**0.293**	**0.297**	**0.334**	**0.378**	**0.378**	**0.320**	**0.096**
Percent foreign born	0.001	-0.020	**-0.099**	**-0.105**	**-0.112**	**-0.081**	**-0.060**	**-0.044**	**-0.041**	-0.025	**-0.070**
Regional dummies											
South Atlantic	**0.151**	**-0.055**	-0.024	**0.052**	**0.114**	**0.170**	**0.176**	**0.196**	**0.212**	**0.224**	**0.128**
East South Central	**0.135**	**-0.049**	-0.018	**0.049**	**0.093**	**0.126**	**0.098**	**0.097**	**0.091**	**0.109**	**0.063**
West South Central	**0.043**	-0.018	-0.016	0.009	0.018	**0.038**	**0.035**	**0.041**	**0.029**	0.026	-0.023
R-square	**0.340**	**0.289**	**0.351**	**0.374**	**0.406**	**0.490**	**0.541**	**0.538**	**0.459**	**0.274**	**0.121**

Note: The number of observations in each model was 3,087. U.S. counties total 3,141, but Alaska and Hawaii were excluded and Virginia independent cities were combined with adjacent counties. Coefficients for rural metro, nonmetro adjacent, and nonmetro nonadjacent dummy variables measure effects relative to the omitted category: urban metro. Coefficients in bold were statistically significant at the .05 level.
Source: USDA, Economic Research Service, using data from the U.S. Census Bureau and the Bureau of Economic Analysis.

equal number of outmigrants. Large birth cohorts tend to have more effective migration patterns (Plane and Rogerson, 1991, Plane, 1992, Pandit, 1997a).

Less efficient migration flows among younger age groups, as well as smaller birth cohorts, likely account for the lower goodness-of-fit values for the cohorts following the baby boomers.

Projecting Population Growth From Migration

These projections were constructed by asking: "What will future migration patterns look like if the most recently measured age-specific migration rates (from the 1990s) stay the same?" Although population projections by metro and nonmetro categories are uncommon, the methodology employed here is similar to that used to project U.S. State populations by the Census Bureau (http://www.census.gov/population/www/projections/projectionsagesex.html). Age affects migration in relatively predictable ways that can be statistically measured. The overall size of age cohorts is also easy to project into the future using forward survival methods because age-specific death rates are relatively fixed (Arias, 2006). Immigration's relatively small impact on older age groups can be measured using the Census Bureau's "best guess" estimates of future, age-specific immigration flows. Thus, projections can be made of the size of future baby boom cohorts in different types of metro and nonmetro counties.

Projections are made of population change from migration for this decade and the next, starting with the 1990s regression models:

$$CMR_{1990-2000,j,a} = \alpha_0 + \beta_{i,a} \left| Var_{1990,i,j} \right| + \varepsilon$$

where β refers to unique regression coefficients for each cohort a and independent variable i measured for each county j. With the exception of employment change (measured for 1990-93) and the ERS Natural Amenities Index (which uses relatively fixed environmental measures), the regression models measure independent variables as of 1990 to model 1990-2000 net migration. To project cohort migration rates for 2000-10 for each county j and age group a, we compiled the same set of county-level independent variables i from the 2000 census and 2000-2003 employment data (app. table 4). Cohort migration rates for 2000-10 were estimated by multiplying the regression coefficients for a given age group a by the updated independent variables:

$$CMR_{2000-2010,j,a} = \alpha_0 + \beta_{i,a} \left| Var_{2000,i,j} \right| + \varepsilon$$

An initial estimate of population change from net migration (NMI) for age group a in county j was calculated as:

$$NMI_{2000-2010,j,a} = \frac{Population_{2000,j,a}}{CMR_{2000-2010,j,a}}$$

A final adjustment was needed to ensure that the sum of all net domestic migration added to zero. This was accomplished in three steps: first, the immigration portion of NMI was subtracted (using Census projections of age-specific immigration); second, a domestic migration weight was calculated (*DMW*) as each county's proportion of total domestic net migration for each age group; and third, a final estimate of population change from net migration (*NM*) was calculated as:

$$NM_{2000-2010,j,a} = NMI_{2000-2010,j,a} - \left(NMI_{2000-2010,j,a} * DMW_{2000-2010,j,a}\right)$$

For the 2010-20 period, the analysis used the estimated cohort migration rates for 2000-10 but applied them to projected 2010 populations. The 2010 populations were calculated by adding the estimates of net migration and the Census Bureau's estimates of immigration, and subtracting expected age-specific deaths.

Example of County-Level Net Migration Projection

The projection procedure is shown here for 50-54 year olds (as of 2000) in hypothetical "Logan County," NH (app. table 4). In 1990, Logan County was a nonmetro, nonadjacent county with a median home value of $87,000 and 17 percent of housing units designated for seasonal or occasional use (column 1). These values were slightly higher than the national average as reflected in the z-scores greater than 0 (column 2), and Logan County's unemployment and employment change was lower than average.

To estimate the 1990-2000 cohort migration rate for Logan County based on the regression results, the individual z-scores were weighted by their respective coefficients (column 4) and then summed. This resulted in a predicted cohort migration rate for 50-54 year olds of 0.0863.

Appendix table 4
An example of cohort migration projection for hypothesized Logan County, New Hampshire

	(1) Actual values (1990)	(2) Z-score	(3) Regression coefficients	(4) 1990-2000 estimation (2*3)	(5) Actual values (2000)	(6) Z-score	(7) 2000-10 estimation (3*6)
Constant	NA	NA	0.0728	0.0728	NA	NA	0.0728
Percent unemployed	0.06	-0.2232	0.0083	-0.0019	0.04	-0.5120	-0.0042
Employment change	0.13	-0.0868	0.0223	-0.0019	0.12	0.3186	0.0071
Median home value	87000	0.9992	-0.0027	-0.0027	94300	0.2080	-0.0006
Natural amenity index	-0.57	-0.2734	0.0424	-0.0116	-0.57	-0.2750	-0.0117
Percent seasonal housing	0.17	1.2270	0.0529	0.0649	0.14	0.8344	0.0442
Percent urban	0.00	-1.2256	-0.0088	0.0108	0.02	-1.2329	0.0109
Population density	37.64	-0.1105	0.0040	-0.0004	40.21	-0.1122	-0.0004
Percent married couples no kids	0.31	-0.3278	0.0456	-0.0149	0.38	0.2400	0.0109
Percent foreign born	0.02	-0.1033	-0.0122	0.0013	0.02	-0.3458	0.0042
Rural metro	0	0	0.0206	0	0	0	0.0000
Adjacent	0	0	0.0015	0	0	0	0.0000
Non-adjacent	1	1	-0.0302	-0.0302	1	1	-0.0302
South Atlantic	0	0	0.0657	0	0	0	0.0000
East South Central	0	0	0.0589	0	0	0	0.0000
West South Central	0	0	0.0159	0	0	0	0.0000
Predicted net migration rate (sum of columns)				**0.0863**			**0.1030**

Note: NA=not applicable. Regression coefficients shown here are unstandardized versions of the coefficients reported in appendix table 3.
Source: USDA, Economic Research Service, using data from the U.S. Census Bureau and the Bureau of Economic Analysis

By 2000, Logan County's median home value had increased to $94,300 but had decreased relative to other counties (columns 5 and 6). Employment change in 2000-03 had decreased slightly from 1990-93 but had increased relative to other counties. Weighting these newer values once again by the coefficients generated in the regression analysis gave us a projected cohort migration rate for the period 2000-10 of 0.103 for 50-54 year olds (as of 2010). The increase in the projected rate from the 1990s to the 2000s was driven mostly by the change in relative home values and employment conditions, as well as an increase in the percent of married couples with no children.

To calculate population change from net migration during 2000-10, the estimated rate of 0.103 was applied to the cohort that would be 50-54 in 2010. The population of 40-44 year olds in Logan County in 2000 was 2,623, indicating that the county had increased its population by 270 in this age cohort. A small adjustment to this estimate was made to ensure that total domestic net migration adds to zero. Population change during 2010-20 was derived by applying the same estimated net migration rate for 50-54 year olds in Logan County to the estimated population for the cohort of 40-44 year olds in 2010.

Projection Ranges

Almost all projections include low, middle, and high alternatives, based on different assumptions about future demographic trends. For instance, the U.S. Census Bureau offers different scenarios for the U.S. population in 2050 depending on potential changes in fertility, mortality, and immigration rates (http://www.census.gov/population/www/projections/usinterimproj/idbsummeth.html). Published statistics usually focus on the middle series projections (http://www.census.gov/population/www/projections/projectionsagesex.html).

For this report, projection alternatives differ in terms of their impact on population change among baby boom cohorts in nonmetro locations. A low projection comes from a straightforward application of 1990-2000 regression results to 2000 Census indicators and 2000-03 employment changes. The findings for baby boomers represent what will happen if they respond to county characteristics in a fashion similar to their predecessors in the 1990s. This is considered a low expectation because, in the past, baby boomers have shown a higher preference for rural and small-town settings, especially where scenic amenities are high and housing prices are low. Employment change during 2000-03 also was quite low, leading to a more conservative projection series.

The high projection series was constructed by increasing the parameter estimates for the ERS Natural Amenities Index, the percent seasonal housing units, and, where significant, the nonmetro location coefficients. It is impossible to predict future employment trends with a high degree of confidence, but current economic conditions lend credence to the decision to calculate a more conservative projection series by maintaining the low employment change conditions of 2000-03. Most of the findings shown in this report are from a middle projection series, which was derived as the average of the high and low series.